

19 janvier 2011

Corneille dans l'ombre de Molière Comment identifier un auteur?



(Jean-Léon Gérôme Une collaboration Corneille et Molière –1874)

Dominique Labbé

(PACTE - Institut d'Etudes Politiques de Grenoble)

Résumé :

Molière a-t-il composé les pièces présentées sous son nom ? Plusieurs indices historiques montrent que ce n'est pas le cas. Au XVII^e siècle on présentait les comédies satiriques des grands auteurs sous le nom de "comédiens poètes". Molière ne s'est pas comporté en écrivain et aucun de ses contemporains ne l'a traité comme tel. Au contraire, de nombreuses rumeurs ont couru, certaines désignant P. Corneille comme l'auteur des pièces présentées par Molière. Cinq indices statistiques confirment ces rumeurs : la distance intertextuelle, les classifications automatiques, les combinaisons des verbes usuels, le sens des mots les plus fréquents, la longueur des phrases.

Abstract :

Several evidences show that Moliere did not write his plays. During the French 17th century, the comedies were presented by comedians – like Molière – and not by the authors who wrote them. Molière did not behave as an author and none, among his contemporaries, considered him as a writer. On the contrary, several testified that Corneille wrote some of these plays. These testimonies are confirmed by several statistical indices: intertextual distances, classifications, combinations of the usual words, meanings of the keywords, lengths of the sentences.

Le tableau de J.-L. Gérôme représente Corneille et Molière, deux célèbres écrivains français du XVII^e siècle.

Dans ce tableau, qui est Corneille et qui est Molière ? Et que sont-ils en train de faire ? Voilà ce que nous allons découvrir.

Pierre Corneille (1606-1684) est l'écrivain français le plus célèbre du XVII^e siècle. Il a écrit 34 pièces – en majorité des tragédies – dont la liste est donnée en annexe 1 de cette conférence. Il est l'auteur des deux plus grands succès théâtraux du siècle : *Le Cid* et *Psyché*.

Molière (1622-1673) était un comédien célèbre qui a présenté les pièces de théâtre les plus jouées durant ces 4 derniers siècles : *Tartuffe*, *Dom Juan*, *Misanthrope*, *Bourgeois gentilhomme*, *Malade imaginaire* (liste en annexe 2). Sa célébrité est telle qu'on dit que le français est la « langue de Molière ».

Mais Molière a-t-il écrit les pièces qu'on lui attribue ? Comment peut-on identifier un auteur ?

Des études historiques et statistiques permettent de répondre scientifiquement à cette question (pour une présentation des méthodes d'attribution d'auteur : Love 2002).

I. QUATRE ETUDES HISTORIQUES

1. Le théâtre français au XVII^e siècle

Cette partie est un résumé. Pour plus de détails, on peut se reporter à notre ouvrage "*Si deux et deux...*" (Labbé 2009).

Durant la seconde moitié du XVII^e, 6 pièces de théâtre sur 10 ont été présentées par des acteurs qui ne les avaient pas écrites. C'est le cas de 9 comédies sur 10. On appelait ces acteurs des « comédiens poètes »¹. Voici les noms des plus célèbres :

Baron, Brécourt, Champmeslé, Dancourt, Desfontaines, Dorimond, Hauteroche, La Thuillerie, Montfleury, Poisson, Raisin, Rosimond, Villiers...

Desfontaines est le plus ancien. Il a terminé sa carrière dans la troupe où Molière a commencé la sienne.

Voici quelques exemples qui permettront de comprendre les raisons de ce système.

Exemple 1 : le comédien poète

Le comédien poète est une comédie jouée par la troupe de Molière, juste après sa mort (1673). Elle montre la manière dont les troupes utilisaient les pièces apportées par les comédiens poètes. Elle a été présentée, puis publiée, sous le nom de Montfleury alors qu'elle avait été écrite par... Thomas Corneille (1625-1705). En effet, d'après le livre de compte de la troupe, des droits d'auteur lui ont été versés. Il est important de noter que :

- la troupe, les gazettes et l'éditeur savaient tous que T. Corneille était le véritable auteur et pourtant, à l'époque, tous ont dit et écrit que c'était Montfleury,

- juste après *Le comédien poète*, la troupe de Molière a joué *La mort d'Achille*, une tragédie de... T. Corneille.

Autrement dit, en 1673, T. Corneille a présenté une tragédie sous son propre nom et a utilisé un homme de paille pour présenter une comédie, avec la complicité générale.

¹ Pour une présentation de ces comédiens et des pièces présentées sous leur nom, voir le site en ligne anglais CESAR (Calendrier Electronique des Spectacles sous l'Ancien Régime et la révolution).

Exemple 2 : une comédie sans titre mais avec deux auteurs

La Comédie sans titre a été le plus grand succès théâtral de la saison 1683-1684. Elle a été présentée par le comédien Raymond Poisson puis publiée dans ses « Œuvres complètes » en 1687. R. Poisson est mort en 1690. En 1694, la comédie sans titre a été republiée avec en sous-titre « revue et corrigée par son véritable auteur » : E. Boursault.



Monsieur Poisson, que je priay de la mettre sous son nom, pour quelques raisons que j'avois, & qui ont cessé, eut assez de scrupule pour ne vouloir estre que l'Econome d'un bien dont je luy avois abandonné la propriété. Quand il eut assuré le succès de cet Ouvrage il cessa d'en vouloir estre l'Auteur : Et le refus qu'il fit d'accepter une réputation qui ne luy appartenoit pas, merite que ma reconnaissance ajoute ce témoignage à celle qu'il s'est acquise.

BOURSULT;

Dans sa préface, Boursault explique qu'il a vendu sa pièce à Poisson et que celui-ci, avant sa mort, a "cessé d'en vouloir être l'auteur" ! En 1683, au moment de la création de la pièce :

- la troupe, les journalistes, l'éditeur savaient que l'auteur était Boursault, tous ont dit que c'était Poisson !

- Quelques semaines après la *Comédie sans titre*, la Comédie française a joué *Mary Stuart*, une tragédie de... Boursault.

Autrement dit, Boursault, comme T. Corneille, présentait ses tragédies sous son nom et ses pièces légères grâce à un prête-nom... avec la complicité des acteurs, des journaux et des éditeurs.

Notre livre donne d'autres exemples.

- Après 1678, T. Corneille était associé avec un autre acteur : Hauteroche. La troupe, les gazettes et les éditeurs connaissaient la vérité et pourtant tout le monde a affirmé que Hauteroche était "auteur" des pièces qu'il présentait.

- Le célèbre fabuliste J. de la Fontaine, a présenté plusieurs pièces de façon anonyme ou sous le nom d'un autre comédien poète : Champmeslé.

Etc.

Les raisons de ce système

Les parisiens aimaient beaucoup les comédies légères et satiriques comme celles de Molière. Plus de la moitié des recettes des troupes provenaient de ces comédies. Mais elles étaient condamnées par l'Eglise, la Cour, l'Académie française, de telle sorte que les auteurs véritables voulaient rester dans l'ombre.

De plus, il n'y avait pas de protection légale du droit d'auteur et pas de statut légal pour les troupes. Pour protéger son exclusivité, la troupe faisait acheter le texte par un acteur. Ce comédien supervisait la création et l'exploitation, comme le font les metteurs en scène aujourd'hui. Il subissait les critiques et, éventuellement, la censure comme cela est arrivé plusieurs fois à Molière...

Aucun des comédiens poètes contemporains de Molière n'était écrivain. Molière était-il une exception ? Ecrivait-il lui-même les comédies qu'il présentait comme on le croit aujourd'hui ?

2. Qui était Molière ?

Il reste une centaine de documents authentiques de l'époque à propos de Molière (Jurgens & Maxfield-Miller 1963). On y découvre que :

- Molière était un riche financier qui se prétendait faussement "valet de chambre du roi" et qui se présentait parfois comme noble.

- Ses deux résidences étaient richement meublées, mais il avait très peu de livres, aucune bibliothèque et pas de bureau.

- Il ne reste aucun manuscrit de lui, à l'exception d'une vingtaine de signatures sur des actes officiels (Dulait 1967). Pas de lettre, aucune note de sa main, aucun témoignage qu'il ait entretenu une correspondance avec un de ses contemporains.

- Entre 1659 et 1673, période pendant laquelle il est supposé avoir écrit tous ses chefs d'œuvre, son emploi du temps est connu grâce au "registre de La Grange" (Young 1997). Cet emploi du temps ne lui permettait pas d'écrire une moyenne de deux pièces par an, comme il est supposé l'avoir fait.

3. Les témoignages des contemporains

Trois éditeurs

Trois éditeurs ont clairement indiqué que Molière n'a pas écrit :

- *Le dépit amoureux* (1662). Dans une dédicace en tête de l'ouvrage, l'éditeur indique que cette pièce est de « l'auteur le plus approuvé de ce siècle ». Tous les spécialistes du XVII^e siècle sont d'accord : en 1662, cette formule désigne P. Corneille.

- *Psyché* (1671). Dans un avertissement au début du livre, l'éditeur indique que les trois quarts des vers ont été écrits par P. Corneille.

- *Dom Juan* (mis en vers par T. Corneille en 1677). La couverture porte le seul nom de Molière, mais dans un avertissement, l'éditeur écrit : "jusqu'à maintenant cette pièce a été *présentée* sous le nom de Molière" (et non pas "composée" ou "écrite" par...)

Trois critiques

Donneau de Visé, fondateur du premier magazine littéraire (*le Mercure galant*) :

« Ses enfants ont plus d'un père (...) Tout le Parnasse s'assemble quand il veut faire quelque chose » (1663).

Robinet du Laurens, éditeur d'une lettre rimée hebdomadaire sur les nouvelles de Paris et de la Cour :

« On ne peut pas dire que Molière soit une source vive mais seulement un bassin qui reçoit ses eaux d'ailleurs pour ne point le traiter plus mal en le comprenant dans la comparaison (...) avec des ânes seulement capables de porter des grands fardeaux » (1663).

"On donne de Corneille sa *Bérénice* sans pareille sur le théâtre de Molière... et son charmant *Bourgeois gentilhomme*" (1670).

Boileau, satiriste et critique littéraire, est souvent cité en faveur de Molière. Pourtant, en 1662, dans un de ses premiers poèmes, il écrit qu'un *Cornelius* est caché derrière Molière comme un *Cornelius* (Scipion Emilien) était caché derrière Térence. En 1665, il sous-entend que Molière ne sait pas faire des vers. Enfin, en 1674, il sous-entend que Molière n'a pas écrit le *Misanthrope* et écrit qu'il est "un faux plaisant, à grossière équivoque qui, pour me divertir, n'a que la saleté".

Aucun des contemporains de Molière ne l'a traité comme un écrivain. Par exemple, La Fontaine l'a rencontré en 1662 et ne l'a plus jamais mentionné dans ses écrits, ni dans sa correspondance. Madame de Sévigné, qui recevait beaucoup à dîner, ne l'a jamais invité, etc.

4. Pourquoi les frères Corneille ?

P. Corneille a donc été désigné à plusieurs reprises comme étant l'auteur de certaines pièces présentées par Molière (notamment le *Dépit amoureux*, *Psyché* et le *Bourgeois gentilhomme*). Pourquoi lui ?

Les deux frères Corneille (Pierre et Thomas) ont écrit les plus grands succès théâtraux du siècle (*Le Cid*, *Psyché*, *Timocrate*, *Circé*). Logiquement, les troupes s'adressaient à eux pour leur fournir des comédies à succès. Les deux frères étaient pauvres et, après la mort de Mazarin, l'arrestation de Fouquet et la mise à l'écart de la Reine mère, ils n'avaient plus de protecteurs. Or, à partir de 1660, ils font face à des dépenses considérables pour l'établissement de leurs 10 enfants. Comme nous l'expliquons dans notre livre, leur production de l'ombre est la seule ressource financière qui a pu permettre ces dépenses.

D'ailleurs, en 1674, Boileau a explicitement accusé les frères Corneille d'être "affamés d'argent" et de faire un "métier mercenaire".

Comment en être sûr ?

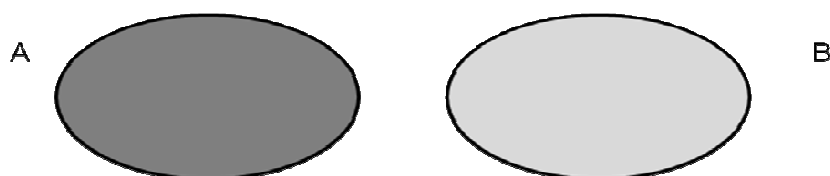
Nous allons présenter une méthode permettant d'attribuer à un auteur connu des textes d'origine douteuse ou inconnue : la distance intertextuelle combinée avec des classifications automatiques. Plusieurs indices viendront confirmer cette attribution.

II. CINQ INDICES STATISTIQUES

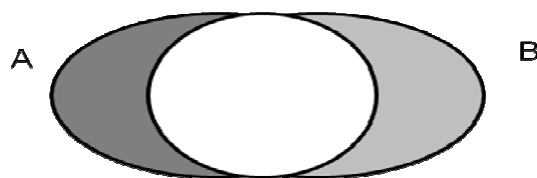
Cette méthode, appliquée à Corneille et Molière, a été présentée pour la première fois dans Labbé et Labbé (2001). Voir également : Labbé et Labbé 2003 et Labbé 2007.

1. La distance intertextuelle

La distance entre deux textes est mesurée comme on mesure la distance séparant deux points dans l'espace. L'unité de mesure est le "mot". Soit deux textes A et B



On les superpose et on compte le nombre de mots différents (zones grisées dans le schéma ci-dessous).



Pour obtenir une mesure normalisée, on calcule la distance relative en rapportant le nombre de mots différents au total du nombre de mots contenu dans A et B .

Cette mesure a les propriétés d'une « distance euclidienne » (longueur du segment de droite unissant deux points). Cet adjectif signifie "conforme à la géométrie d'Euclide" (par un point il ne passe qu'une parallèle à une droite située hors de ce point). Les propriétés d'une distance euclidienne sont : l'identité (la distance d'un point à lui-même est nulle), la symétrie (le résultat est le même que l'on mesure AB ou BA) et l'inégalité triangulaire (le chemin direct entre A et B est toujours plus court qu'en passant par un point C non situé sur le segment AB).

Les résultats obtenus avec la formule (2) varient uniformément entre un minimum de zéro — distance nulle : mêmes vocables et mêmes fréquences relatives — et un maximum de 1 : tous les mots sont différents. Elle peut aussi être exprimée en %, ‰, etc. Cette mesure relative permet de comparer les résultats obtenus sur un grand nombre de textes de longueurs différentes.

La distance intertextuelle est influencée par 4 facteurs principaux, soit par ordre d'importance décroissante :

- le genre : oral et écrit, prose, vers, comédie et tragédie...
- l'époque où a été rédigé le texte car chaque époque a un vocabulaire particulier,
- l'auteur
- le thème (personnages, lieux, principaux motifs)

Dès lors, pour déterminer l'auteur d'un texte d'origine douteuse ou inconnue, il suffit de le confronter à d'autres — dont l'origine n'est pas douteuse — écrits dans un même genre et à la même époque. NB : le théâtre doit être comparé au théâtre et, autant que possible, les tragédies entre elles, les comédies entre elles, etc.

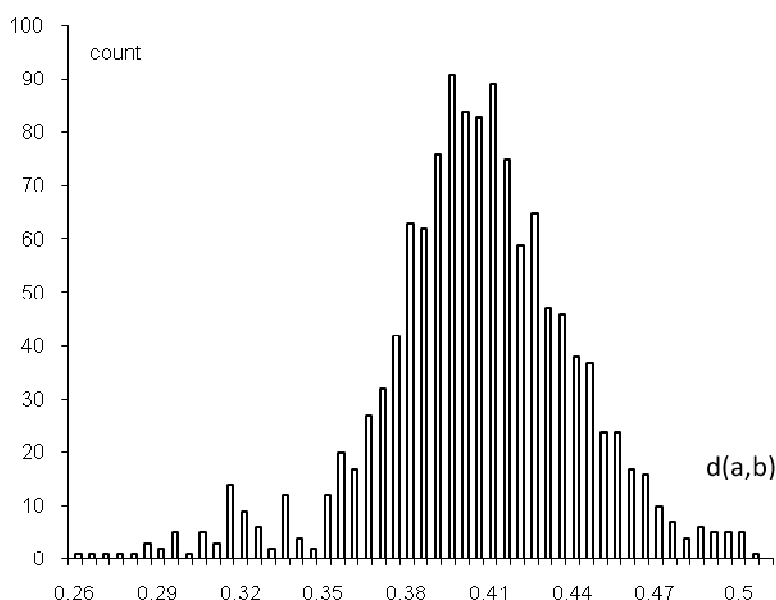
La mise au point de la méthode

La méthode d'attribution d'auteur a été mise au point grâce à un très grand nombre d'expériences. En voici un exemple.

A la demande de Gerard Ledger et de Thomas Merriam, une série d'expériences en aveugle ont été réalisées sous leur contrôle. Ils ont relu le compte rendu et en ont accepté la publication (Labbé 2007). Lors de la première expérience, G. Ledger a soumis 52 textes anglais anonymés – le "corpus Oxquarry" – en demandant lesquels étaient écrits par les mêmes auteurs. Il avait choisi certains textes parce qu'il semblait difficile de les distinguer bien qu'ayant été écrits par des auteurs différents. Plusieurs indications ont été fournies : les 52 textes sont d'un même genre et à peu près contemporains – il s'agit de romans de la seconde moitié du XIXe – il y a plusieurs auteurs et chacun de ces auteurs a au moins deux textes mais pas forcément du même ouvrage.

La formule (2) appliquée à chaque paire de textes donne 1326 distances différentes $(52 \times 51)/2$ qui ont été rangées par valeurs croissantes dans des classes d'intervalles égaux.

Histogramme des distances rangées par valeurs croissantes et intervalles de classe égaux (corpus Oxquarry)



Les valeurs se distribuent selon la courbe en cloche ("Laplace-Gauss"). Mais il y a des anomalies aux deux extrémités (distances plus grandes ou plus petites que la moyenne \pm deux écarts types. On formule deux hypothèses :

- les 70 paires de textes "trop" proches par rapport à la moyenne du corpus sont de mêmes auteurs,
- les 26 distances "trop" élevées par rapport à la moyenne séparent des textes d'auteurs différents.

Ces 96 couples (soit 7% du total) permettent d'attribuer 47 textes sur 52 (tableau ci-dessous). 5 textes ne peuvent être « mariés » à aucun autre : même auteur et thèmes ou époque très éloignés ou textes contemporains sur des thèmes proches par deux auteurs différents ?

Détermination des auteurs grâce aux distances remarquables

N°	Textes	Auteurs
1	1D , 2W, 1K, 2R, 1N, 2E, 1F	Stevenson
2	1S, 2M, 2I, 1C	Morris
3	1E, 1Q, 1B, 2A	Butler
4	2L, 2T, 2J, 2X, 1L, 2D, 1X, 2F	Hardy (Well beloved, Wessex, Judde)
5	1H, 2Z, 2S	Hardy (Madding)
6	2O, 2P, 2V	Forster
7	1Z, 2C, 1Y	Tressel
8	1P, 2U, 1R	Chesterton
9	2H, 1G, 1O	Conrad (Lord Jim)
10	2N, 2Q, 2K	Conrad (Almayer)
11	1M, 2Y, 1W, 2G	Orczy
12	1J, 2B	Morris

Après l'expérience, T. Merriam et G. Ledger ont livré les noms des auteurs (dernière colonne à droite) : il n'y a aucune erreur mais deux auteurs (Hardy et Conrad) figurent dans deux groupes pour des livres différents). Les deux hypothèses sont donc vérifiées (les distances les plus faibles désignent un même auteur et les distances les plus fortes des auteurs différents).

De nombreuses expériences comme celle-ci ont été réalisées. Jusqu'à ce jour, elles ont toutes validé ces deux hypothèses. Par exemple, en 2006 avec Denis Monière, nous avons révélé l'identité du journaliste qui avait servi de "plume de l'ombre" pour deux Premiers ministres québécois et nous avons identifié, sans erreur, les discours qui avaient été rédigés par ce collaborateur (Labbé & Monière 2006, Labbé & Monière 2008).

De nombreux chercheurs ont utilisé ces méthodes avec succès. Par exemple, sur les politiciens anglais (Arnold 2005), les pièces de théâtre élisabéthaines (Merriam, 2002, 2003a, 2003b, 2005), les premiers ministres italiens (Tuzzi et Al. 2010).

Application à Corneille et Molière

Appliquée au théâtre du XVII^e (Corneille, Molière, Racine, Quinault et Mairêt), cette même méthode classe correctement la plupart des textes, mais elle apporte deux surprises :

- deux comédies en alexandrins de P. Corneille - *le Menteur* (1642) et *la suite du Menteur* (1643) – sont classées avec 14 comédies présentées par Molière entre 1659 et 1673 : douze comédies en alexandrins et deux comédies en prose (*Dom Juan* et *l'Avare*). Toutes les distances séparant ces pièces sont remarquablement faibles (annexe 3),

- deux comédies "sérieuses" présentées par Molière, également en alexandrins, (*Dom Garvie* et *Psyché*) sont classées avec les tragédies et les comédies héroïques contemporaines de P. Corneille (annexe 4).

On peut ajouter trois choses.

Premièrement, la distance intertextuelle a été testée sur des dizaines de milliers de textes, sans jamais rencontrer de tels croisements entre deux œuvres d'auteurs différents, sauf... quand les deux auteurs ne font qu'un, comme dans le cas de Gary et Ajar (Labbé 2004a et Lafon et Peters 2006), ou quand la même plume de l'ombre a travaillé pour deux employeurs différents (Monière et Labbé, 2006).

Deuxièmement, le XVII^e siècle fournit plusieurs contre-épreuves intéressantes. Par exemple, Corneille et Racine ont chacun écrit une tragédie sur l'amour impossible entre un empereur romain (Titus) et une reine orientale (Bérénice). Ils l'ont fait au même moment, en aveugle, en alexandrins et en respectant les fameuses règles de la création théâtrale. Et tous deux avaient en tête la liaison entre Louis XIV et sa belle-sœur Henriette d'Angleterre. Le lieu de l'action, les personnages, les thèmes étaient donc identiques (d'où un vocabulaire commun important...). La distance entre ces deux pièces (0,256 ou 2 560 mots différents pour 10 000) est plus élevée que toutes celles constatées entre les pièces en vers de Molière et les deux *Menteurs* de Corneille ou entre *Dom Garcie* et les tragédies de Corneille, alors que toutes ces pièces sont séparées par un laps de temps important et que les thèmes sont très divers (annexes 3 à 6).

Le XVII^e fournit d'autres exemples. Les *Sophonisbe* de Mairet et de Corneille se distinguent bien alors qu'ils ont travaillé dans le même genre, sur le même thème, les mêmes événements, les mêmes personnages, en suivant la même trame narrative. Les comédies de Quinault, ou celle de Racine (les *Plaideurs*) se distinguent bien de celles de Corneille et de Molière qui sont impossibles à départager (voir dernière colonne de l'annexe 3). Quand Quinault, Corneille et Molière travaillent ensemble sur *Psyché*, Quinault se distingue des deux autres... qui sont impossibles à départager.

Enfin, troisième réponse, les grands tableaux de chiffres sont difficiles à manipuler manuellement. Il y a un risque d'erreur ou d'omission. Des automates peuvent faire ce travail plus efficacement et offrir des représentations graphiques plus simples à consulter que les tableaux de chiffres.

2. Les classifications

L'objectif est double :

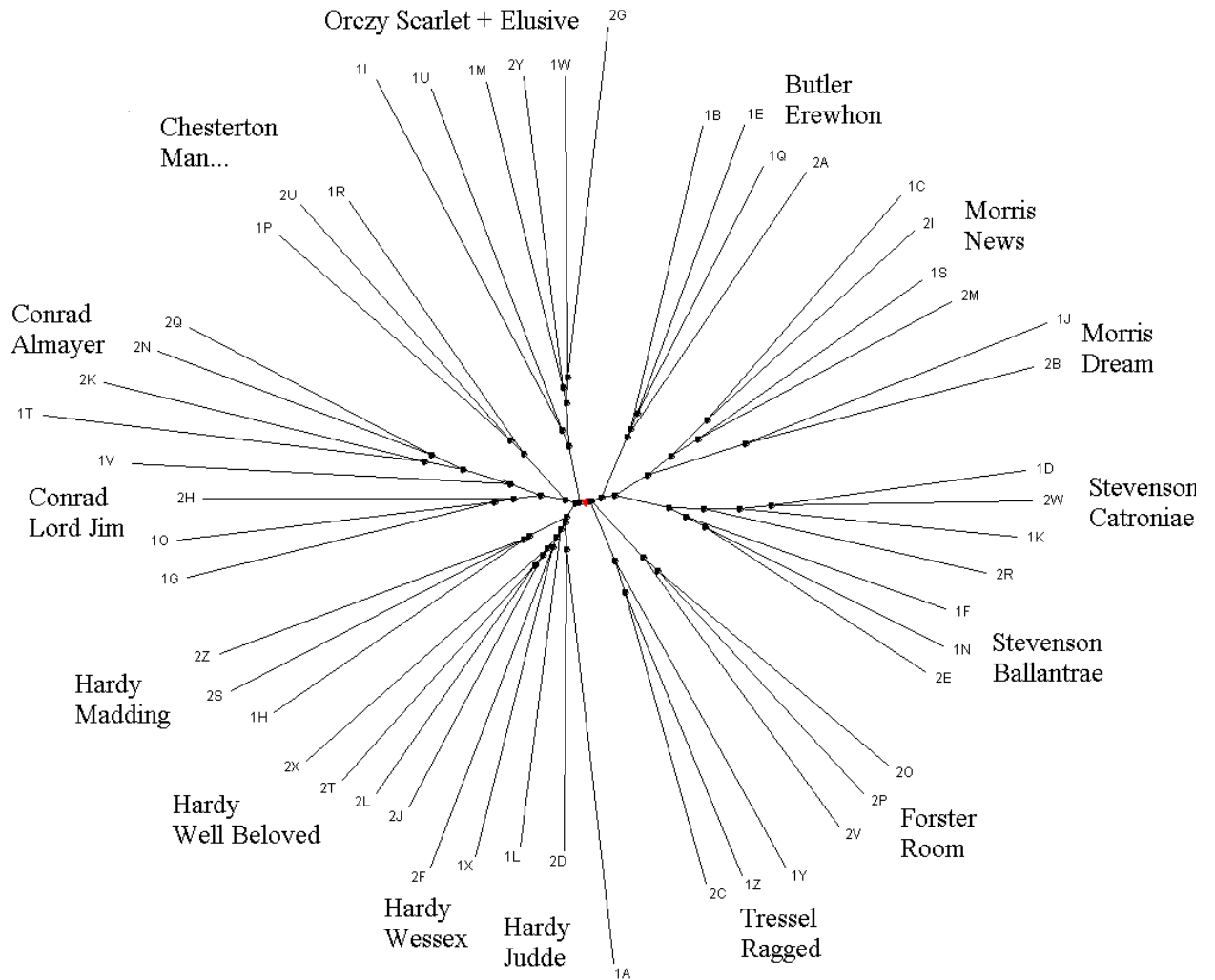
- rechercher les meilleurs groupements possibles sans intervention humaine. Deux critères sont utilisés : d'une part, les distances entre les individus composant un groupe doivent être les plus courtes possibles ; d'autre part, les distances séparant les différents groupes ainsi constitués, doivent être les plus grandes possibles ;
- offrir la meilleure représentation possible, en deux dimensions, de cet ordre alors que celui-ci se déploie dans un espace à beaucoup de dimensions (1 324 dans le cas de l'expérience en aveugle ci-dessus).

Nous utilisons trois méthodes : la classification hiérarchique sur les deux premières composantes d'une analyse en composantes principales (avec le logiciel R) ; les classifications hiérarchique et arborée sur les distances intertextuelles à l'aide de logiciels développés en collaboration avec X. Luong (1988), M. Rulhman (2003) et C. Labbé (Labbé & Labbé 2006, 2008).

La classification arborée est très utilisée en biologie et en génétique. Elle repose sur la propriété suivante : si toutes les distances séparant les individus étudiés sont euclidiennes, il existe un "arbre" qui représente exactement les positions respectives de ces individus les uns par rapport aux autres et les meilleurs groupements possibles.

Appliquée à l'expérience en aveugle sur les romanciers anglais du XIX^e, l'algorithme donne la figure ci-dessous.

Classification arborée du corpus Oxquarry



NB : les noms des auteurs et des œuvres ont été ajoutés après l'expérience.

Dans cette figure, les **feuilles** terminales figurent chacune un texte ; les **nœuds** intermédiaires indiquent les meilleurs groupements possibles. La distance entre deux points est figurée par le **chemin** unissant ces points et la longueur de ce chemin est proportionnelle à la distance originelle correspondante.

Attention : la position de chaque feuille dans l'espace est sans importance. Seules comptent les proximités, les contrastes et la longueur des chemins à parcourir pour aller d'un texte à un autre.

Dans la figure ci-dessus, tous les textes sont correctement attribués à leurs auteurs et aux œuvres respectives. Combien de combinaisons différentes peut-on réaliser avec 56 objets. En théorie, 56 ! (factorielle de 56), c'est-à-dire : $56 * 55 * 54 \dots * 3 * 2 = 7,11 \text{ e}74$. Même en tenant compte des propriétés de symétrie et d'identité déjà mentionnées, on obtient encore un nombre de combinaisons comportant une trentaine de zéros... Il n'y a donc aucune chance pour que ce genre d'épreuve puisse être réussie par hasard.

Appliquée aux pièces de Corneille et de Molière, la classification arborée confirme les quatre anomalies signalées ci-dessus : les deux *Menteurs* de P. Corneille sont classés au milieu des pièces en vers présentées par Molière ; *Psyché* et *Dom Garcie* sont rattachées au reste de

l'œuvre de P. Corneille (annexe 6). On remarquera que ce graphique a été réalisé par X. Luong qui ignorait le titre et les auteurs des textes qu'il classait.

3. Trois autres indices

Trois autres indices statistiques conduisent aux mêmes conclusions.

Les combinaisons de mots préférées.

Les combinaisons des verbes les plus fréquents chez Corneille et Molière révèlent les mêmes proximités (Annexe 7).

Racine ne partage avec Corneille et Molière que trois combinaisons (soulignées dans le tableau en annexe) : "pouvoir voir", "pouvoir faire" et "pouvoir être", mais avec un classement et des densités très différentes. En revanche, Corneille et Molière en ont cinq en commun (en gras) dont les trois premières dans le même ordre et avec des densités voisines (en italiques). Étant donné le nombre des combinaisons possibles, la probabilité pour qu'une telle "coïncidence" survienne au hasard est infinitésimale.

Il existe un seul cas comparable dans les 4 derniers siècles de littérature française : Gary et Ajar². Depuis huit ans, personne n'a pu trouver un autre exemple concernant deux auteurs réellement différents...

Le sens des mots usuels.

Autre indice : le sens spécifique que chaque auteur donne aux principaux mots qu'il emploie. L'étude des réseaux sémantiques révèle que, chez Corneille et Molière, les principaux vocables ont le même sens, ou plutôt, que ceux de Molière forment un sous-ensemble dans ceux de Corneille. Le plus évocateur est le mot "amour" – substantif le plus employé par Corneille (Labbé & Labbé 2006). Là encore, ces significations sont propres à Corneille et ne se retrouvent pas chez ses contemporains. Cette démonstration a été également faite, en utilisant les « collocations » lors d'un congrès à Louvain-la-Neuve (Labbé 2004b). La démonstration est en ligne depuis 7 ans et elle n'est pas contredite à ce jour.

Longueur de phrase

Dans un genre donné, à une époque précise, la plupart des auteurs se singularisent par des longueurs de phrases différentes. Cela offre un outil auxiliaire pour l'attribution d'auteur (Labbé & Labbé, 2010). *Dans le théâtre du XVIIe, il existe deux singularités qui concernent... Corneille-Molière :*

Première exception : la distribution des longueurs de phrases dans le *Menteur* et la *Suite du Menteur* (P. Corneille) ne diverge pas de celle observée dans les onze comédies en alexandrins présentées par Molière. Cela est vrai pour chacun des 24 couples de pièces considérés séparément.

Deuxième exception : la distribution des longueurs de phrases dans *Dom Garvie* (Molière) et dans *Psyché* (Corneille présenté sous le nom de Molière) ne diverge pas significativement de celle observée dans les tragédies (et les tragi-comédies) en alexandrins de P. Corneille.

² Annexe 8. Voir également les combinaisons préférées d'une « plume de l'ombre » qui a travaillé pour deux Premiers ministres québécois (Monière et Labbé 2006).

De nouveau, Corneille et Molière utilisent des longueurs de phrases semblables. Ici ce sont 18 pièces de Corneille qui sont concernées (de *Rodogune* à *Suréna*).

Conclusions

Tous nos travaux sont consultables sur le site "archives en ligne" du CNRS (HAL-SHS). Nos programmes et nos données sont dans le domaine public. La plupart des documents historiques cités sont en ligne ou publiés. Tout est vérifiable, tout est reproductible.

Revenons maintenant au tableau de Gérôme. L'auteur l'a baptisé "Une collaboration". Il représente Corneille – pauvrement vêtu - lisant une pièce à Molière qui vient en prendre livraison. Notez que ce tableau a été peint en 1874 : depuis longtemps on sait qui a écrit les chefs d'œuvre présentés par Molière ! Au moins P. Louys³, H. Poulaille ou H. Wouters l'ont écrit avant nous.

A Paris, durant la seconde moitié du XVIIe, 6 pièces de théâtre sur 10 – et 9 comédies sur 10 - ont été présentées par des comédiens poètes – comme Molière - et non par les écrivains qui les avaient composées. Un très grand nombre d'indices historiques et statistiques montrent que Corneille et Molière ont fait de même. Corneille a écrit :

- toutes les pièces en vers présentées par Molière (*l'Etourdi*, *le Dépit amoureux*, *Dom Garcie*, *l'Ecole des maris*, *les Facheux*, *l'Ecole des femmes*, *la Princesse d'Elide*, *Tartuffe*, *le Misanthrope*, *Amphytrion*, *les Femmes savantes*),

- en prose : *Dom Juan* et *l'Avaro*, et au moins 4 des 5 actes du *Bourgeois gentilhomme* et du *Malade imaginaire*.

Il faut maintenant redécouvrir l'œuvre de Corneille et l'histoire réelle du théâtre français du XVIIe siècle. En effet, P. Corneille n'est pas le seul concerné : il reste à identifier une bonne partie de l'œuvre de son frère Thomas, de Jean de la Fontaine et de quelques autres écrivains... La moitié des pièces produites à cette époque attendent de trouver leur véritable auteur !

Nous savons donc maintenant comment reconnaître l'auteur d'un texte douteux ou d'origine inconnue. La distance intertextuelle combinée avec la classification arborée constitue l'outil principal. Les autres indices – comme les combinaisons des verbes fréquents, le sens des mots les plus usuels ou la longueur des phrases – offrent d'utiles compléments. Cela va permettre de résoudre les grandes énigmes que l'histoire littéraire nous a léguées.

Enfin, ce petit exposé suggère combien l'étude de la littérature et de la langue pourrait être profondément renouvelée par la statistique appliquée et l'informatique.

Remerciements

Cette recherche est collective. P. Hubert, C. Labbé et D. Monière ont collaboré à la mise au point de la distance intertextuelle ; C. Labbé, X. Luong et M. Ruhlman à celle de la classification arborée. Beaucoup d'autres chercheurs nous ont apporté leur aide – comme J.-G. Bergeron, M. Brugidou, J. Jolissaint, G. Ledger, J. et N. Leselbaum, Tom Merriam, J. Savoy et beaucoup d'autres.

³ Les articles de Pierre Louys sont reproduits en annexe de Boissier 2004.

Références

Tous nos travaux cités sont accessibles en ligne à partir de notre page personnelle ou sur le site « Archives ouvertes » du CNRS. On consultera également le site corneille-moliere.org.

- Arnold Edward (2005). "Le discours de Tony Blair (1997-2004)". *Corpus*, 4, p. 55-77.
- Baudeau de Sommaize (1660). "Préface". *Les véritable précieuses*. Reproduit dans : Mongrédien Georges. *Comédies et pamphlets sur Molière*. Paris : Nizet, 1986, tome I, p. 35-67.
- Boileau Nicolas (1662). "Stances à M. Molière" in *Les délices de la poésie galante*. Reproduit dans *Œuvres complètes*. Paris : Gallimard, 1966, p. 246.
- Boileau Nicolas (1665). "Satire II". Reproduit dans *Œuvres complètes*. Paris : Gallimard, 1966.
- Boileau Nicolas (1674). "L'art poétique". Reproduit dans *Œuvres complètes*. Paris : Gallimard, 1966.
- Boissier Denis (2004). *L'affaire Molière*. Paris : Jean-Cyrille Godefroy.
- Donneau de Vizé (1663). *Réponse à l'Impromptu de Versailles*. Reproduit dans : Mongrédien Georges. *Comédies et pamphlets sur Molière*. Paris : Nizet, 1986, tome II.
- Duchêne Roger (1998), *Molière*, Paris, Fayard.
- Dulait Suzanne (1967). *Inventaire raisonné des autographes de Molière*. Genève : Droz, 1967
- Jurgens Madeleine & Elisabeth Maxfield-Miller (1963). *Cent ans de recherches sur Molière*. Paris : Imprimerie nationale.
- La Fontaine Jean de (1968). *Œuvres diverses*. Paris : Gallimard, La Pléiade.
- Labbé Cyril & Labbé Dominique (2001b). "Inter-Textual Distance and Authorship Attribution Corneille and Molière". *Journal of Quantitative Linguistics*. 8-3, December 2001, p. 213-231.
- Labbé Cyril & Labbé Dominique (2003). "La distance intertextuelle". *Corpus*. 2, p. 95-118.
- Labbé Cyril & Labbé Dominique (2005). "How to Measure the Meanings of Words ? Amour in Corneille's Work". *Language Resources Evaluation*. 39, p. 335-351.
- Labbé Cyril & Labbé Dominique (2006). "A Tool for Literary Studies. Intertextual Distance and Tree Classification". *Literary and Linguistic Computing*. 21-3, p. 311-326.
- Labbé Cyril & Labbé Dominique (2007). *Corneille a écrit 16 pièces représentées sous le nom de Molière. Réponses à : Vîprey Jean-Marie et Ledoux Claude-Nicolas, 'About Labbé's "Inter-textual Distance"'*. Grenoble : PACTE-IEP.
- Labbé Cyril & Labbé Dominique (2008). "Peut-on se fier aux arbres ?". In Heiden Serge et Pincemin Bénédicte (Eds). *9^e Journées internationales d'analyse statistique des données textuelles (Lyon, 12-14 mars 2008)*. Lyon : Presses universitaires de Lyon, 2008, volume 2, p. 635-645.
- Labbé Cyril & Labbé Dominique (2009). "Existe-t-il un genre épistolaire ? Hugo, Flaubert et Maupassant". *Xe journées de l'Erla*. Brest novembre 2009.
- Labbé Cyril & Labbé Dominique (2010). "Ce que disent leurs phrases". In Bolasco Sergio, Chiari Isabella, Giuliano Luca (Eds). *Proceedings of 10th International Conference Statistical Analysis of Textual Data*. Rome : Edizioni Universitarie di Lettere Economia Diritto, 2010, Vol 1, p. 297-307.
- Labbé Dominique (2004a). *Romain Gary et Emile Ajar*. Grenoble : Cerat-IEP, mai 2004.
- Labbé Dominique (2004b). *Corneille et Molière. Table ronde 7e Journées d'Analyse des Données Textuelles*. Louvain-la-Neuve 11 mars 2004. Grenoble : CERAT-IEP.
- Labbé Dominique (2007). "Experiments on Authorship Attribution by Intertextual Distance in English". *Journal of Quantitative Linguistics*, April 2007, 14-1. p. 33-80.
- Labbé Dominique (2009). *Qui a écrit Tartuffe ?* Montréal : Monière et Wollank. Réédition : *Si deux et deux sont quatre Molière n'a pas écrit Don Juan*. Paris : Max Milo.
- Labbé Dominique (2010). "Ce que disent les phrases de Corneille et Molière". Communication devant les Xe Journées Internationales d'Analyse des Données Textuelles. Rome : juin 2010.

- Labbé Dominique & Monière Denis (2006). "L'influence des plumes de l'ombre sur les discours des politiciens". In Condé Claude et Viprey Jean-Marie. *Actes des 8e Journées internationales d'Analyse des données textuelles*. Besançon, II, 687-696.
- Labbé Dominique & Monière Denis (2008). *Les mots qui nous gouvernent. Le discours des premiers ministres québécois : 1960-2005*. Montréal : Monière-Wollanck.
- Lafon Michel & Peeters Benoît (2006). *Nous est un autre*. Paris, Flammarion.
- Love Harold (2002). *Attributing Authorship: An Introduction*. Cambridge : Cambridge University Press.
- Luong Xuan (1988). *Méthodes d'analyse arborée. Algorithmes, applications*. Thèse pour le doctorat ès sciences. Paris : Université de Paris V.
- Merriam Thomas (2002). "Intertextual Distances between Shakespeare Plays, with Special Reference to *Henry V* (verse)". *Journal of Quantitative Linguistics*. 9-3, 260-273.
- Merriam Thomas (2003a). "An Application of Authorship Attribution by Intertextual Distance in English". *Corpus*. 2, 167-182.
- Merriam Thomas (2003b). "Intertextual Distances, Three Authors". *Literary and Linguistic Computing*, 18-4, 379-388.
- Merriam Thomas (2005). *The Identity of Shakespeare in Henry VIII*. The Renaissance Institute, Tokyo.
- Poulaille Henry (1957). *Corneille sous le masque de Molière*. Paris : Grasset, 1957.
- Reynier Gustave (1892). *Thomas Corneille, sa vie et son théâtre*. Paris : Hachette.
- Robinet Charles (1663). *Panegyrique de l'Ecole des femmes*. Reproduit dans : Mongrédien Georges. *Comédies et pamphlets sur Molière*. Paris : Nizet, 1986, tome I.
- Robinet Charles (1670). *Gazette rimée*. Reproduit dans Brooks William. "Le théâtre et l'opéra vus par les gazetiers Robinet et Laurent". *Papers on French Seventeenth Century Literature*. Tome XVII, 1993, 76.
- Ruhlmann Mathieu (2003). *Analyse arborée. Représentation arborée par la méthode des groupements*. Rapport de stage sous la direction de Labbé Cyril et Labbé Dominique. Grenoble : Polytech'Grenoble et Cerat-IEP, août 2003.
- Tuzzi Arjuna, Popescu Ioan-Iovitz & Altmann Gabriel (2010). "Quantitative Analysis of Italian Texts". *Studies in Quantitative Linguistics*, 6.
- Wouters Hippolyte & Ville de Goyet Christine de (1990). *Molière ou l'auteur imaginaire ?* Bruxelles : Complexe.
- Young Bert E. & Grace P. (1997). *Le registre de La Grange*, Genève, Slatkine.

Annexe I.
Les pièces de Pierre Corneille.

		Création	Genre	Longueur en mots
1	Mélite	1630 ?	Comédie	16 690
2	Clitandre	1631	Tragi-comédie	14 402
3	La Veuve	1631	Comédie	17 661
4	La Galerie du Palais	1632	Comédie	16 140
5	La Suivante	1633	Comédie	15 160
6	Comédie des Tuileries	1634	Comédie	3 627
7	Médée	1635	Tragédie	14 269
8	La Place Royale	1634	Comédie	13 801
9	L'illusion comique	1636	Comédie	15 428
10	Le Cid	1636	Tragi-comédie	16 677
11	Cinna	1639	Tragédie	16 126
12	Horace	1640	Tragédie	16 482
13	Polyeucte	1641	Tragédie	16 472
14	Pompée	1642	Tragédie	16 492
15	Le menteur 1	1642	Comédie	16 653
16	Le menteur 2	1643	Comédie	17 675
17	Rodogune	1644	Tragédie	16 842
18	Théodore	1645	Tragédie	17 121
19	Héraclius	1647	Tragédie	17 433
20	Andromède	1650	Tragédie	15 514
21	Don Sanche	1650	Comédie héroïque	16 947
22	Nicomède	1651	Tragédie	16 923
23	Pertharite	1651	Tragédie	17 121
24	Oedipe	1659	Tragédie	18 618
25	Toison d'Or	1661	Tragédie	20 343
26	Sertorius	1662	Tragédie	17 675
27	Sophonisbe	1663	Tragédie	16 858
28	Othon	1664	Tragédie	16 971
29	Agésilas	1666	Tragédie	18 227
30	Atilla	1667	Tragédie	16 788
31	Tite et Bérénice	1670	Comédie héroïque	16 697
32	Pulchérie	1672	Tragédie	16 630
33	Suréna	1674	Tragédie	16 545

34	Psyché Corneille	1671	Comédie en vers	10 067
35	Psyché Molière	1671	Comédie en vers	4 816
36	Psyché Quinault	1671	Comédie en vers	1 399

Sources : Charles Marty-Laveaux. *Œuvres complètes de P. Corneille*. Paris : Hachette 1862.
Collection Les Grands écrivains de la France.

Annexe 2

L'œuvre théâtrale représentée sous le nom de Molière

		Création	Genre	Longueur mots
37	La jalousie	Avant 1659	Comédie prose	3 501
38	Médecin volant	Avant 1659	Comédie prose	3 876
39	L'étourdi*	1659	Comédie vers	18 671
40	Dépit amoureux*	1659	Comédie vers	16 242
41	Précieuses ridicules	1660	Comédie prose	6 648
42	Sganarelle*	1660	Comédie vers	6 042
43	Dom Garcie*	1661	Comédie héroïque vers	17 049
44	L'école des maris*	1661	Comédie vers	10 536
45	Les fâcheux*	1661	Comédie vers	7 922
46	L'école des femmes*	1662	Comédie vers	16 625
47	Critique de l'école	1663	Comédie prose	8 610
48	L'impromptu	1663	Comédie prose	7 168
49	Mariage forcé	1664	Comédie prose	6 058
50	Princesse d'Elide*	1664	Comédie vers et prose	11 333
51	Le Tartuffe*	1664	Comédie vers	18 271
52	Dom Juan*	1665	Comédie prose	17 452
53	L'amour médecin	1665	Comédie prose	6 147
54	Le Misanthrope*	1666	Comédie vers	17 180
55	Médecin malgré lui	1666	Comédie prose	9 317
56	Mélicerte*	1666	Comédie vers	5 540
**	Comédie pastorale	1667	Comédie vers libres	732
57	Le sicilien	1667	Comédie prose	5 375
58	Amphytrion*	1668	Comédie vers libres	15 117
59	Georges Dandin	1668	Comédie prose	11 009
60	L'avare*	1668	Comédie prose	21 033
61	M. de Pourceaugnac	1669	Comédie prose	11 803
62	Amants magnifiques*	1670	Comédie vers & prose	11 983
63	Bourgeois gentilhom.*	1670	Comédie prose	17 132
64	Fourberies de Scapin	1671	Comédie prose	14 245
65	Escarbagnas	1671	Comédie prose	5 564
66	Femmes savantes*	1672	Comédie vers	16 863
67	Malade imaginaire*	1673	Comédie prose	19 919

* Pièce écrite, en tout ou partie, par P. Corneille

** Pièce retirée dans les expériences à cause de sa petite taille.

Sources : Eugène Despois. *Œuvres complètes de Molière*. Paris : Hachette, 1876. Collection Les Grands écrivains de la France.

Annexe 3
Distances séparant les deux *Menteurs* (Corneille) et les *Plaideurs* (Racine)
de toutes les pièces de Molière

N°	Pièces	Genre	Le Menteur (Corneille 1642)	Suite du Menteur (Corneille 1643)	Les Plaideurs (Racine : 1668)
15	Le Menteur (1642)	Vers	0,0000	0,1797	0,2961
16	La suite du Menteur (1643)	Vers	0,1797	0,000	0,2933
34	Psyché Corneille (1671)	Vers	0,2875	0,2730	0,3647
36	Psyché Molière (1671)	Vers	0,3294	0,3249	0,3777
37	La jalousie du barbouillé (avant 1660)	Prose	0,3411	0,3312	0,3269
38	Médecin volant (avant 1660)	Prose	0,3097	0,2932	0,3021
39	L'étourdi (1658)	Vers	0,2048	0,2060	0,2691
40	Dépit amoureux (1658)	Vers	0,2154	0,2114	0,2702
41	Précieuses ridicules (1660)	Prose	0,3143	0,3138	0,3139
42	Sganarelle ou le cocu imagin. (1660)	Vers	0,2585	0,2527	0,2928
43	Dom Garcie de Navarre (1661)	Vers	0,2802	0,2728	0,3585
44	L'école des maris (1661)	Vers	0,2231	0,2168	0,2791
45	Les fâcheux (1661)	Vers	0,2477	0,2476	0,3062
46	L'école des femmes (1662)	Vers	0,2256	0,2167	0,2608
47	Critique de l'école des femmes (1663)	Prose	0,3229	0,3187	0,3405
48	L'impromptu de Versailles (1663)	Prose	0,3212	0,3157	0,3233
49	Mariage forcé (1664)	Prose	0,3217	0,3024	0,3196
50	Princesse d'Elide (1664)	Vers Prose	0,2518	0,2426	0,3135
51	Le Tartuffe (1664)	Vers	0,2416	0,2315	0,2753
52	Dom Juan (1665)	Prose	0,2619	0,2471	0,2812
53	L'amour médecin (1665)	Prose	0,2917	0,2888	0,2871
54	Le Misanthrope (1666)	Vers	0,2524	0,2331	0,2824
55	Médecin malgré lui (1666)	Prose	0,2970	0,2886	0,2963
56	Mélicerte (1666)	Vers	0,2569	0,2499	0,3217
57	Le sicilien ou l'amour peintre (1667)	Prose	0,2772	0,2598	0,3011
58	Amphytrion (1668)	Vers libres	0,2525	0,2563	0,2966
59	Georges Dandin (1668)	Prose	0,2916	0,2788	0,2922
60	L'Avare (1668)	Prose	0,2566	0,2439	0,2696
61	M. de Pourceaugnac (1669)	Prose	0,2916	0,2820	0,2852
62	Amants magnifiques (1670)	Prose	0,2821	0,2793	0,3295
63	Bourgeois gentilhomme (1670)	Prose	0,2940	0,2816	0,2864
64	Fourberies de Scapin (1671)	Prose	0,2692	0,2626	0,2805
65	Comtesse d'Escarbagnas (1671)	Prose	0,3111	0,3001	0,3048
66	Femmes savantes (1672)	Vers	0,2598	0,2486	0,2829
67	Malade imaginaire (1672)	Prose	0,2861	0,2789	0,2782
<i>Moyenne oeuvre de Molière</i>			<i>0,2761</i>	<i>0,2680</i>	<i>0,3002</i>
<i>Moyenne pièces en vers de Molière</i>			0,2386	<i>0,2315</i>	<i>0,2850</i>
<i>Moyenne oeuvre de Corneille</i>			<i>0,2513</i>	<i>0,2480</i>	<i>0,3532</i>
<i>Moyenne oeuvre de Racine</i>			<i>0,3140</i>	<i>0,3129</i>	<i>0,3763</i>

Annexe 4.

Distances séparant *Dom Garcie* (Molière) et *Psyché* (Corneille et Molière) des dernières pièces de Corneille.

Ultimes pièces de Corneille	Dom Garcie (Molière,1661)	Psyché (Corneille sous le nom de Molière 1671)
Rodogune (1644)	0,2448	0,2310
Theodore (1645)	0,2341	0,2445
Heraclius (1647)	0,2480	0,2729
Andromède (1650)	0,2412	0,2180
DonSanche (1650)	0,2242	0,2514
Nicomède (1651)	0,2445	0,2644
Pertharite (1651)	0,2346	0,2632
Edipe (1659)	0,2234	0,2264
Toison d'or (1661)	0,2212	0,2198
Sertorius (1662)	0,229	0,2378
Sophonisbe (1663)	0,2276	0,2357
Othon (1664)	0,2346	0,2399
Agesilas (1666)	0,2342	0,2327
Attila (1667)	0,2349	0,2270
Tite et Bérénice (1670)	0,2275	0,2347
Psyché (1671)	0,2300	—
Pulcherie (1672)	0,2300	0,2260
Surena (1674)	0,2165	0,2236
<i>Moyenne Corneille</i>	<i>0,2431</i>	<i>0,2435</i>
<i>Moyenne Molière</i>	<i>0,2862</i>	<i>0,2974</i>

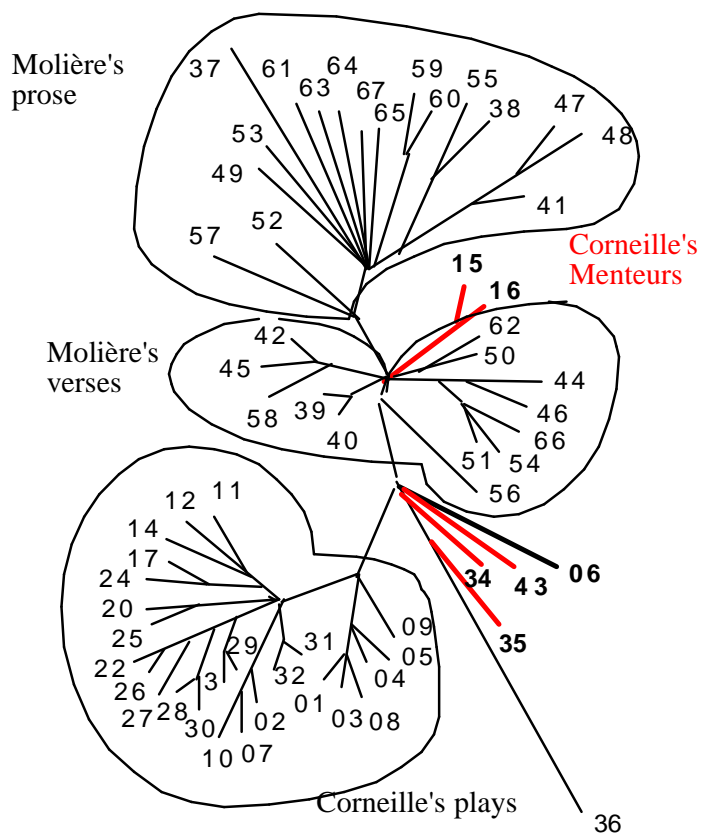
Annexe 5

Principales distances caractéristiques entre Corneille et Racine à l'époque de *Tite et Bérénice*.

	Tite et Bérénice (Corneille, 1670)	Bérénice (Racine, 1670)
CORNEILLE :		
Agésilas (1666)	0.1585	0.2780
Attila (1667)	0.1801	0.2892
Tite et Bérénice (1670)	0	0.2561
Pulchérie (1672)	0.1535	0.2712
Suréna (1672)	0.1559	0.2643
RACINE :		
Andromaque (1667)	0.2587	0.2266
Britannicus (1669)	0.2507	0.2095
Bérénice (1670)	0.2561	-
Bazajet (1672)	0.2620	0.2195
Mithridate (1673)	0.2487	0.2061

Annexe 6

Classification arborée sur l'ensemble des œuvres théâtrales de Corneille et Molière



Ce graphe a été tracé par M. X. Luong de l'université de Nice.

Pour les numéros des pièces, se reporter aux annexes 1 et 2.

Les traits en gras correspondent à : n° 06 Corneille : Comédies des Tuileries (écrite par Corneille pour Richelieu en 1634) ; n° 15 et 16 Corneille : *Le Menteur* et la *Suite du Menteur* (1642 et 1643) ; n° 34 : passages de *Psyché* écrits par Corneille ; n° 35 : passages de *Psyché* attribués par l'éditeur à Molière ; n° 36 : prologue de *Psyché* écrit par Quinault ; n° 43 : *Dom Garvie* présenté par Molière

Annexe 7

Les combinaisons "pseudo-auxiliaire + infinitif" chez Corneille, Molière et Racine
(fréquence pour 100.000 mots)

P. Corneille		Molière		Racine	
Syntagmes	F	Syntagmes	F	Syntagmes	F
<i>faire voir</i>	33,8	<i>faire voir</i>	31,5	aller voir	12,0
<i><u>pouvoir être</u></i>	18,8	<i><u>pouvoir être</u></i>	25,5	<u>pouvoir voir</u>	9,6
<i><u>pouvoir faire</u></i>	18,4	<i><u>pouvoir faire</u></i>	25,5	faire entendre	9,0
faire naître	13,9	vouloir dire	24,9	<u>pouvoir faire</u>	8,4
<i><u>pouvoir voir</u></i>	13,4	<i><u>vouloir faire</u></i>	19,5	aller chercher	7,8
devoir être	12,7	pouvoir dire	14,5	faire parler	7,8
pouvoir souffrir	10,8	pouvoir avoir	13,7	<u>pouvoir être</u>	7,8
<i><u>vouloir faire</u></i>	9,9	aller faire	13,2	venir chercher	7,2
faire connaître	9,6	avoir faire	13,2	faire éclater	6,6
devoir faire	8,7	<i><u>pouvoir voir</u></i>	12,3	falloir partir	6,6

Annexe 8

Les principaux groupes verbaux chez Ajar et Gary (fréquence pour 10.000 mots)

Ajar	F	Gary	F
vouloir dire	7,20	vouloir dire	5,80
pouvoir être	4,90	pouvoir être	3,62
pouvoir faire	4,40	pouvoir faire	2,97
devoir être	3,10	devoir être	2,39
aller faire	2,90	aller faire	2,20
pouvoir vivre	2,70	devoir avoir	2,13
laisser tomber	2,30	pouvoir dire	2,13
devoir avoir	2,10	laisser tomber	1,87
aller voir	2,00	laisser aller	1,55
devoir faire	1,80	faire passer	1,49